ERA5再分析降水数据在中国的适用性分析

刘婷婷[1,2]， 朱秀芳[1,2,3]， 郭 锐[1,2]， 徐 昆[1,2]， 张世喆[1,2]

（1. 北京师范大学环境演变与自然灾害教育部重点实验室，北京 100875；2. 北京师范大学地理科学学部遥感科学与工程研究院，北京 100875；3. 北京师范大学遥感科学国家重点实验室，北京 100875）

摘 要：为了探讨ERA5再分析降水数据在中国区的适用性，以全国728个站点的日降水数据为参考，使用Pearson相关系数、均方根误差、平均绝对误差、探测率、误报率以及公正先兆评分分析了ERA5再分析降水数据在不同时间尺度（月、季）、不同气候区、不同海拔梯度下的精度以及ERA5再分析数据对暴雨和干旱事件的刻画能力。结果表明：ERA5降水数据对日降水事件的识别能力在空间和时间上均存在差异，整体来看：在北温带的精度最高；在夏、秋两季较冬、春两季的精度低；海拔>500 m地区的精度低于海拔≤500 m地区的精度。ERA5数据在进行暴雨识别时，与站点观测数据的偏差较大，且阈值越大（即暴雨越强）偏差越大；基于ERA5计算的不同尺度的标准化降水指数的精度不同，其中3个月尺度的标准化降水指数的精度最高。在进行干旱事件的识别时，阈值越低（即干旱越严重）误差越大。本研究可为ERA5降水数据的使用范围和使用方法提供参考，有助于分析相关研究的不确定性。

关 键 词：ERA5数据；降水；暴雨；干旱

文章编号：

降水量是至关重要的气候变量，与水资源、农业生产、经济发展等密切相关。了解降水的数量、频率、空间分布、变化趋势等信息对于合理利用水资源，制定农业发展策略等有重要意义。然而，很多偏远地区或者经济发展落后地区气象台站稀少甚至根本没有气象台站[1-4]。利用有限的站点降水数据进行空间插值得到空间化的降水数据往往存在很大的误差。为此，不少学者使用大气再分析的降水资料来进行相关研究，例如分析降水的时空特征（印度河流域气温、降水、蒸发及干旱变化特征）[5-7]、提取暴雨洪涝事件、驱动作物模型[8-9]、水文模型（多源降水数据的小流域水文模拟效用评估）[10-13]和陆面模型等[14]。然而，大气再分析数据是数值预报产品和观测资料融合的产物，预报产品的误差、观测数据的误差和同化方法的误差都会影响再分析气候数据的质量[15]。再分析数据产品的精度会直接影响上述基于大气再分析降水资料进行的相关研究中，相关研究结果的不确定性。在气候变化的大背景下，利用再分析数据驱动各类模型模拟气候变化的影响研究与日俱增，验证和评估再分析数据的精度既必要又紧迫[16]。

ERA5是继ERA-Interim后欧洲中期天气预报中心（ECWMF）推出的第五代再分析产品，提供了大量的海洋气候和每小时的气候变量。这些数据以0.25°×0.25°的网格覆盖地球，数据集中包含200多个参数，提供了大量的逐小时大气、陆地和海洋气候变量。该数据基于改进的三维变分技术，拥有时空分辨率高、更新快、参数多等优点，受到了人们的广泛关注。相关研究指出ERA5相对于ECWMF的第四代再分析产品ERA-Interim有了很大的提升[17-19]。例如，Wang等[20]比较了ERA-Interim和ERA5在北极地区的2 m气温、降雪和总降水数据，发现相较于ERA-Interim，ERA5在夏季和秋季的降水精度有所提升；Betts等[21]使用加拿大萨斯喀彻温省4个气象站的每

收稿日期：2021-03-21； **修订日期**：2021-07-07
基金项目：国家重点研发计划（2019YFAO606900）；国家自然科学基金面上基金（42077436）资助
作者简介：刘婷婷（1998-），女，在读硕士，主要从事极端气候研究. E-mail: 202021051185@mail.bnu.edu.cn
通讯作者：朱秀芳（1982-），女，副教授，博士，主要从事农业遥感、自然灾害研究. E-mail: zhuxiufang@bnu.edu.cn

小时气候站数据对ERA5和ERA-Interim的地表温度、风速、降水以及长、短波向下辐射通量进行了评价，结果表明，与ERA-Interim相比，ERA5的地表温度数据质量有大幅度提升，但风速数据质量较差，降水和长、短波向下辐射通量与ERA-Interim相差不大；Beck等[22]在空间分辨率为0.1°的条件下评估了美国ERA5的日降水量数据，发现与ERA-Interim相比，ERA5再分析降水数据有很大的改善。但是，在以对流风暴为主的地区，ERA5数据有较大误差。

不少研究者在不同地区对该数据集的适宜性进行了评估。例如，Nogueira[23]使用全球降水气候项目2.3版本数据集对月尺度的ERA5降水量进行了验证，发现在热带大部分地区ERA5的偏差和误差通常低于ERA-Interim。但是，ERA5高估了大多数热带海洋和喜马拉雅山的降水。Xu等[24]在阿西尼博因河盆地对包括ERA5在内的6种降水数据进行了评价，发现ERA5大大高估了阿西尼博因河流域的春季降水。Amjad等[25]使用2014—2018年土耳其256个地面气象站点数据对ERA5的降水数据精度进行了评价，结果发现ERA5高估了相对湿润和复杂地形区域的降水，湿偏度达0.5 mm·d^{-1}。

总体来看，已有的ERA5降水数据的适宜性分析往往关注的是再分析产品的数值和实际观测值之间的整体相关程度和偏离情况[22,24,26]，少有对极端气候事件(如暴雨)的模拟精度分析。在少有的针对极端气候事件的分析中，又往往选择典型的气象灾害事件进行评估，不够全面。例如Jiang等[27]使用中国气象局的站点数据分析了ERA5降水数据在6次台风事件中的极端降水模拟的准确性。

基于以上考虑，本文以国家气象科学数据中心的全国范围内728个站点的日降水数据为参考，通过3个方面的分析，来全面评价ERA5降水再分析数据在中国区的适宜性，具体包括：(1) 对比分析ERA5再分析降水数据在不同时间尺度(月、季)上的精度；(2) 对比分析ERA5再分析降水数据在不同气候区和海拔的精度；(3) 分析ERA5再分析降水数据对极端气候事件(暴雨和干旱)的检测能力。该研究可以辅助研究者决定如何使用该数据集、客观评价基于该数据开展的相关研究的不确定性，也为其他的再分析数据适宜性评价研究提供方法参考。

1 数据与方法

1.1 数据

本文使用的数据包括：(1) 来自欧洲中期天气预报中心(ECMWF)ERA5的0.25°×0.25°逐小时降水数据；(2) 来自中国气象局气象信息中心的中国国家级地面气象站降水日值数据集(V1.0)中的共728个站点的1979—2019年的日降水量数据；(3) 来自于资源环境科学与数据中心的中国气候区划图(图1)。本文使用的站点数据在中热带和南热带无站点，因此将这2个区域去除，对剩余8个气候区划进行分析。

1.2 研究方法

根据各气象站点的经纬度，确定各站点所在的ERA5格点，然后提取出这些格点的日降水量数值，并计算出月总降水量和标准化降水指数(SPI)，进而利用6个指标对ERA5再分析降水数据进行评估。这些指标可以分成两类：(1) 用来评价ERA5再分析降水数据与观测降水数据之间的相关性和差异的指标，具体包括Pearson相关系数(r)、均方根误差(RMSE)和平均绝对误差(MAE)。其中，r用来反映站点实测值与ERA5再分析数据的一致性程度；RMSE用来反映ERA5降水序列整体误差水平和波动情况；MAE用来反映ERA5再分析降水数据与站点实测数据的平均绝对偏差程度。(2) 用来评价ERA5对降水、暴雨和干旱事件的捕捉能力的指标，具体包括探测率(POD)、误报率(FAR)以及公正先兆评分(ETS)。其中，POD表示ERA5正确捕捉到的概率，即在所有观测得降水事件中被ERA5正确检测到的降水事件的比例；FAR表示ERA5出现错误的概率，即在所有ERA检测的降水事件中被错误检测出的降水事件的比例；ETS则反映了不同时空上ERA5对降水综合探测的准确性。这些指标的计算公式及最优值见表1。

具体分析过程如下：(1) 首先利用上述6个指标分析ERA5再分析降水数据(日降水量)在中国区的整体精度，其中r、RMSE和MAE用来分析2个数据集之间的整体一致性和差异性，而POD、FAR以及ETS用来分析ERA5再分析数据得到的降水事件的可靠性。其中降水事件的判断标准为日降水量超过0.1 mm·d^{-1}[25]。(2) 接着分析ERA5再分析降水

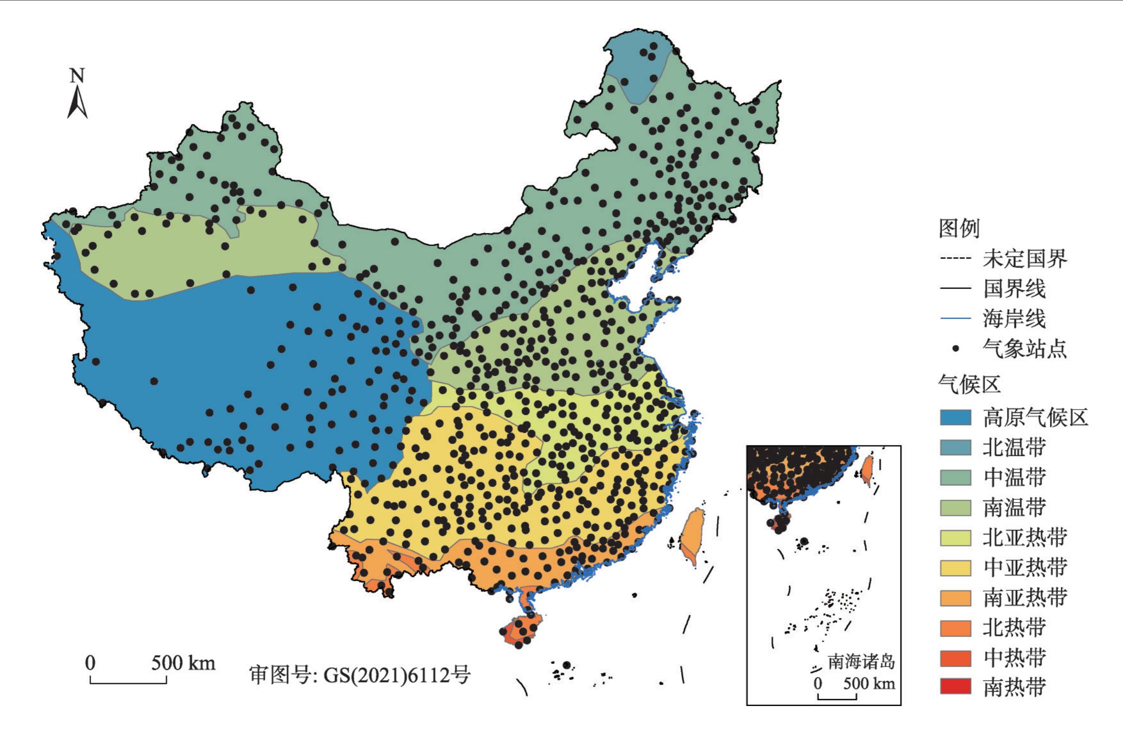


图1 气候区与气象站点分布

Fig. 1 Distributions of climatic areas and meteorological stations

表1 用于评价ERA5数据的指标计算公式

Tab. 1 Calculation formula of indicators used to evaluate ERA5 data

统计指标	计算公式	取值范围	最优值
Pearson 相关系数(r)	$r=\frac{\sum_{i=1}^{N}(E_i-\bar{E})(G_i-\bar{G})}{\sqrt{\sum_{i=1}^{N}(E_i-\bar{E})^2\sum_{i=1}^{N}(G_i-\bar{G})^2}}$	[−1, 1]	1
均方根误差(RMSE)	$\mathrm{RMSE}=\sqrt{\frac{\sum_{i=1}^{N}(E_i-G_i)^2}{N}}$	[0, +∞]	0
平均绝对误差(MAE)	$\mathrm{MAE}=\frac{\sum_{i=1}^{N}\lvert E_i-G_i\rvert}{N}$	[0, +∞]	0
探测率(POD)	$\mathrm{POD}=\frac{\mathrm{hit}}{\mathrm{hit}+\mathrm{miss}}$	[0, 1]	1
误报率(FAR)	$\mathrm{FAR}=\frac{\mathrm{flase\ alarm}}{\mathrm{hit}+\mathrm{false\ alarm}}$	[0, 1]	0
公正先兆评分(ETS)	$\mathrm{ETS}=\frac{\mathrm{hit}-\mathrm{He}}{\mathrm{hit}+\mathrm{miss}+\mathrm{false\ alarm}-\mathrm{He}}$ $\mathrm{He}=\frac{(\mathrm{hit}+\mathrm{false\ alarm})(\mathrm{hit}+\mathrm{miss})}{N}$	[−1/3, 1]	1

注:式中 N 为样本数; E_i 为ERA5的降水量值; $\bar{E}$ 为ERA5降水量的平均值; G_i 为站点的实测降水量值; $\bar{G}$ 为该站点实测降水量的平均值; hit为ERA5数据和实测数据同时检测到事件的次数; miss为实测数据检测到而ERA5数据未检测到事件的次数; false alarm为实测数据未检测到而ERA5检测到事件的次数;He为计算时得到的中间变量。

数据和观测降水数据在不同月份、不同季节、不同气候区以及不同海拔下的 r、RMSE和MAE。其中春、夏、秋、冬四季分别为3—5月、6—8月、9—11月、12月—次年2月;海拔梯度设置为0~200 m、200~500 m、>500 m 3个区间,各海拔区间内的站点数量分别为594、63和71个。(3) 然后利用POD、FAR和ETS分析ERA5降水数据对暴雨事件的识别能力。参考中国气象局对于暴雨的定义,将暴雨分为暴雨(降水量≥50 mm·d^{-1})和大暴雨(降水量≥100 mm·d^{-1})两类。(4) 最后用 r、RMSE和MAE分析基于

ERA5计算出的不同时间尺度(1个月,3个月,6个月和12个月)的SPI的精度,并以3个月时间尺度的SPI(SPI3)为例,参考《气象干旱等级GB/T20481-2006》,使用不同阈值进行不同等级干旱的识别(表2),进而利用POD、FAR和ETS指数评估ERA5对干旱事件的刻画能力。

表2　标准化降水指数(SPI)干旱等级划分

Tab. 2　Drought grade division based on standardized precipitation index (SPI)

类型	标准化降水指数(SPI)
无旱	SPI>-0.5
轻旱	-1.0<SPI≤-0.5
中旱	-1.5<SPI≤-1.0
重旱	-2.0<SPI≤-1.5
特旱	SPI≤-2.0

2　结果与分析

2.1　ERA5降水数据的整体精度分析

图2为POD、FAR、ETS、r、RMSE和MAE 6个评价指标的空间分布图。结果显示,ERA5的日降水精度在空间上存在差异。POD、ETS、r越大且FAR越小的区域为降水数据精度越高的区域。全国有83.6%的站点的POD>0.80,51.9%的站点ETS>0.25,53.3%的站点r>0.50,46.9%的站点FAR<0.50,41.5%的站点RMSE<6,39.6%的站点MAE<2。POD的高值区主要分布在中国南部,包括高原气候区、中亚热带、南亚热带、北亚热带、中热带和南热带,表明ERA5能很好地检测出这些区域的降水。而北温带、中温带和南温带的FAR值较高,表明ERA5会高估这些区域的日降水事件。ETS和r取值在中国东部高于西部,综合来看,ERA5对中国东部降水事件的反映能力强于对中国西部降水事件的反映能力。RMSE和MAE的取值在较为湿润的中亚热带和南亚热带较高,这和该地区降水量绝对值较高有关。

2.2　不同月份和季节的ERA5降水数据精度分析

图3显示的是ERA5数据的日降水量在不同月份和不同季节的精度。不同月份日降水量的r、RMSE和MAE的均值分别为0.482~0.542、2.232~8.530 mm·d^{-1}和0.945~3.438 mm·d^{-1};不同季节日降水量的r、RMSE和MAE的均值分别为0.488~0.525、2.466~8.071 mm·d^{-1}和1.013~3.140 mm·d^{-1}。总体来说,日降水量各季节精度差异较大,夏、秋两季的精度较冬、春两季的精度低。

2.3　不同气候区和海拔的ERA5降水数据精度分析

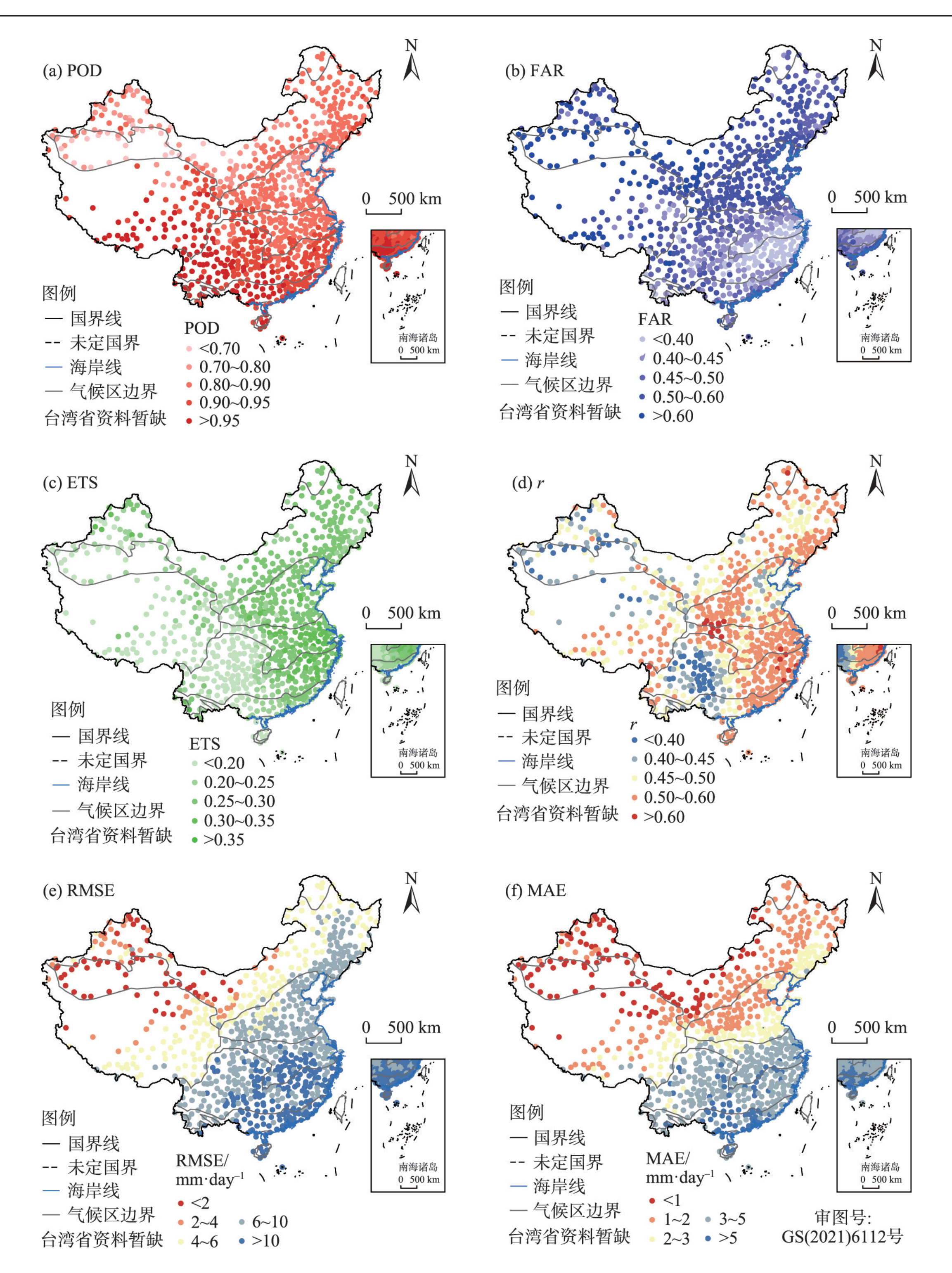
图4为不同气候区和不同海拔梯度下的精度验证结果。不同气候区日降水量的r、RMSE和MAE的均值分别为0.437~0.587、1.825~11.746 mm·d^{-1}和0.619~4.913 mm·d^{-1}。8个气候带中,精度最高的是北温带,其r、RMSE和MAE分别为0.587、4.040 mm·d^{-1}和1.472 mm·d^{-1},中温带和中亚热带的r的范围较大,而南亚热带和北亚热带的RMSE和MAE较大。3个海拔梯度区间中,精度最低的是海拔>500 m的站点,其r、RMSE和MAE分别为0.489、7.505 mm·d^{-1}和3.188 mm·d^{-1};海拔0~200 m的站点的r较高(均值为0.497),但RMSE和MAE值较大;200~500 m的站点的RMSE和MAE最小,分别为3.875 mm·d^{-1}和1.663 mm·d^{-1}。

2.4　暴雨事件精度验证

ERA5数据进行暴雨识别的精度评价结果见图5。其中,日降水量≥50 mm和≥100 mm且发生天数≥10 d的站点分别为626个和377个,主要分布在中国东南部。ERA5的暴雨识别能力在空间上存在差异。当使用50 mm阈值来识别暴雨时,全国有73.6%的站点POD>0.1,51.9%的站点FAR>0.75,53.7%的站点ETS>0.1。POD高值主要集中分布在中国的东部地区(中温带南部、南温带、北亚热带、中亚热带东部和南亚热带东部),FAR高值集中分布在中温带东部、南温带西部和中亚热带西部,ETS高值主要分布在北亚热带、中亚热带东部和南亚热带东部,这说明ERA5数据能较好地捕捉中国东部地区的暴雨事件,但在中温带东部、南温带西部和中亚热带西部的误报率较高,整体来看对东南部(北亚热带、中亚热带东部和南亚热带东部)的暴雨事件捕捉的最好。当使用100 mm阈值来识别暴雨事件时,POD更低、FAR更高、ETS更低,全国仅有23.6%的站点POD>0.1,72.9%的站点FAR>0.75,12.7%的站点ETS>0.1,说明对≥100 mm的暴雨识别能力更差,但空间分布两者较为相似。

2.5　干旱事件精度验证

图6是利用ERA5降水数据计算出的1月尺度的SPI(SPI1)和3月尺度的SPI(SPI3)的验证结果。图7是利用ERA5降水数据计算出的6月尺度的SPI



注:POD为探测率,FAR为误报率,ETS为公平先兆评分,r为Pearson相关系数,RMSE为均方根误差,MAE为平均绝对误差。下同。

图2 ERA5降水数据整体精度

Fig. 2 Overall accuracy of ERA5 precipitation data

(SPI6)和12月尺度的SPI(SPI12)的验证结果。4种不同时间尺度的SPI指示不同时间长度的干旱,SPI1主要反映1个月尺度的干旱,SPI3反映季度上的干旱,SPI6表征的是半年尺度的干旱,而SPI12则表示年尺度上的干旱。由于表征的尺度不同,其精度也有所不同。整体来看,4个尺度的SPI的相关系数、RMSE和MAE的空间分布规律均较为相似:r在青藏地区较低,在东部地区较高;RMSE和MAE在东南地区较高,在青藏地区较低。SPI1、SPI3、SPI6和SPI12的r的平均值分别为0.765、0.746、0.723和

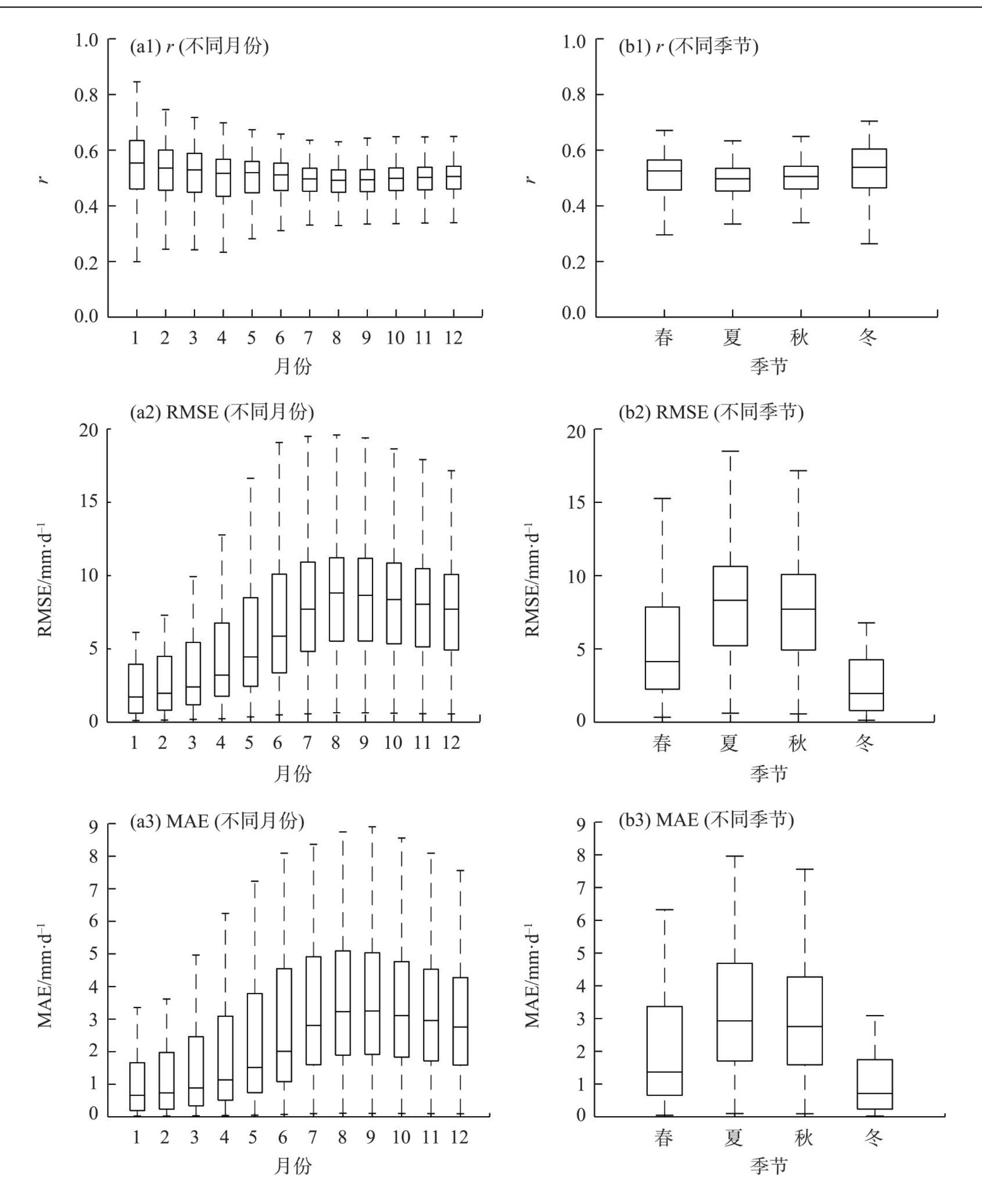


图3　不同月份和季节的ERA5降水数据精度

Fig. 3　Accuracy of ERA5 precipitation data in different months and seasons

0.686，RMSE的平均值分别为0.657、0.701、0.733和0.774，MAE的平均值分别为0.701、0.543、0.576和0.612。SPI1和SPI3的*r*较高且RMSE较低，SPI6和SPI12的*r*较低且RMSE较高，SPI3的MAE最低，SPI1的MAE最高。因此，综合来看，使用ERA5降水数据计算出的SPI3的精度最高。

图8是阈值为−0.5和−1.0的干旱的识别精度结果。图9是阈值为−1.5和−2.0的干旱的识别精度结果。阈值为−0.5、−1.0、−1.5和−2.0识别出的干旱事件历时超过10个月的站点数量分别为728、725、719和509，随着阈值的减小（即干旱强度的增强），识别出的干旱事件历时超过10个月的站点数量减少。ERA5识别出的阈值为−0.5、−1.0、−1.5和−2.0的干旱事件的POD的均值分别为0.680、0.588、0.494和0.426，FAR的均值分别为0.310、0.407、0.511和0.561，ETS的均值分别为0.381、0.349、0.298和0.270，可以看出阈值越小（即干旱强度越强）时，ERA5的干旱识别效果越差。4种阈值识别的干旱的POD、FAR和ETS的空间分布较为相似。以阈值−0.5为例，表3统计了各气候区POD、FAR和ETS

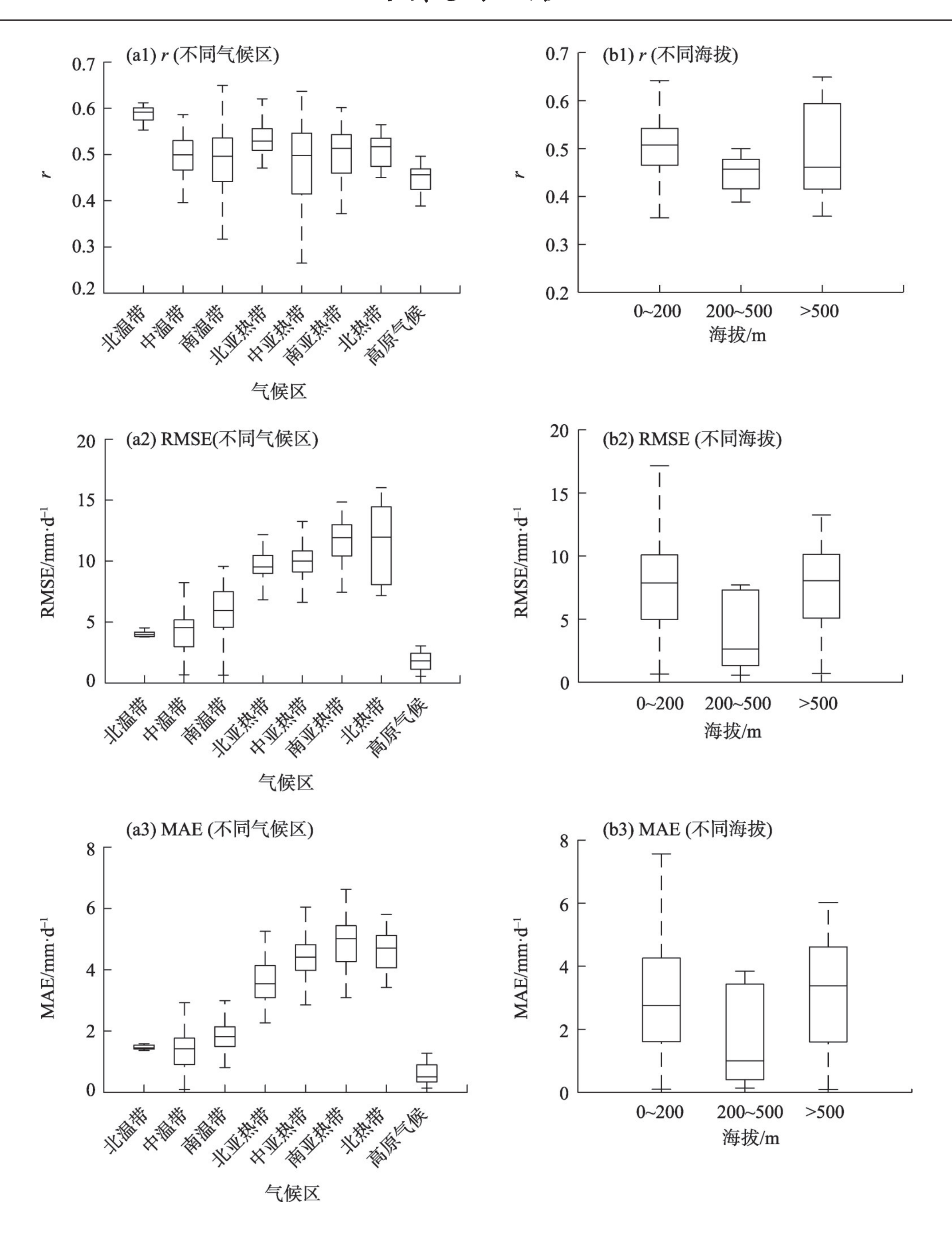


图4　不同气候区和海拔梯度下ERA5降水数据精度

Fig. 4　Accuracy of ERA5 precipitation data in different climate zones and elevation gradients

的均值，整体来说ERA5数据对北温带、南温带和北亚热带的干旱捕捉能力较好，在高原气候带、中亚热带和北热带更容易出现误报，其中北温带的POD最高、FAR最低、ETS最高，干旱识别效果最好，高原气候区的POD最低、FAR最高、ETS最低，干旱的识别效果最差。

3　讨 论

ERA5是欧洲中期天气预报中心的最新一代的再分析产品，其降水量数据可以为分析降水的时空特征、提取暴雨洪涝事件、驱动作物模型、水文模型和陆面模型等提供输入数据。本研究以728个站点的日降水量数据为参考，首次分析了ERA5数据在中国区的精度。相比以往分析最大的不同是关注了ERA5对极端气候事件（暴雨和干旱）的刻画精度。伴随着气候变化，极端气候事件频发，给自然和人类社会带来了众多负面影响和经济损失。评估极端气候条件下的各种影响，发展有效应对气候

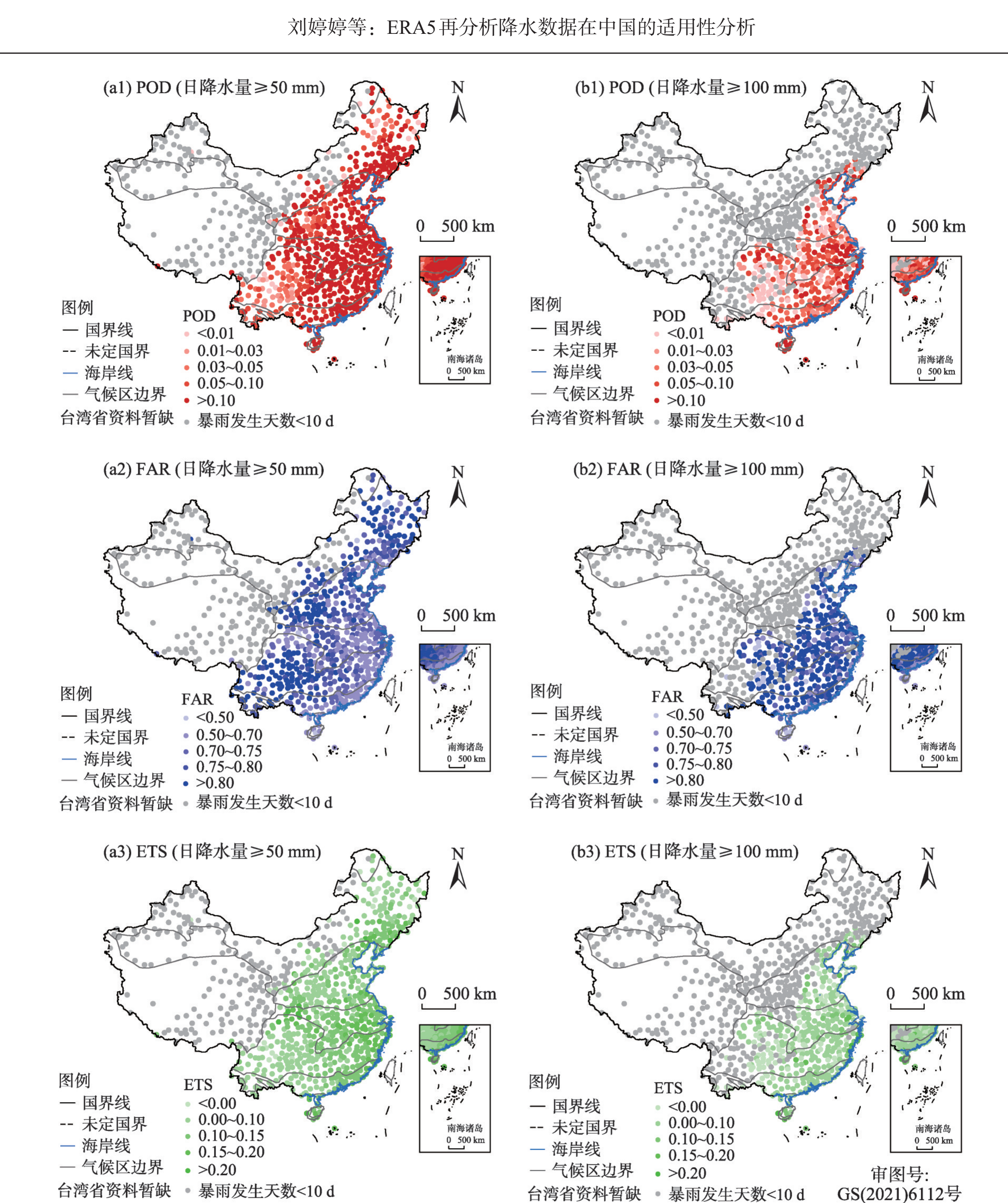


图5 ERA5识别暴雨事件的精度

Fig. 5 Identification accuracy of heavy rain event based on ERA5

变化的适宜性策略，都有赖于可靠的能反映极端气候的数据产品。然而，以往相关气候数据产品的精度分析忽略了对极端事件的捕捉和刻画能力。本文弥补了这一不足，拓展了验证气候数据集的角度，为类似研究提供了方法参考。此外，本文研究发现ERA5数据能够较好地描述3个月尺度的干旱(SPI3)，但干旱强度越强，干旱识别结果越差。总体来说，ERA5对于中国东南部的暴雨的刻画能力最强。这些结论为相关研究考虑是否选择使用ERA5数据，选择使用哪个区域的ERA5数据，选择基于ERA5数据分析什么内容等提供了参考依据。

本文研究发现ERA5数据的精度在不同气候区、不同季节、不同海拔梯度表现出一定的差异。相较于西部，ERA5再分析降水对于东部降水的模拟效果更好，这与成晓裕等对再分析降水数据的研究结果[28]类似；相较于冬、春两季，夏、秋两季的

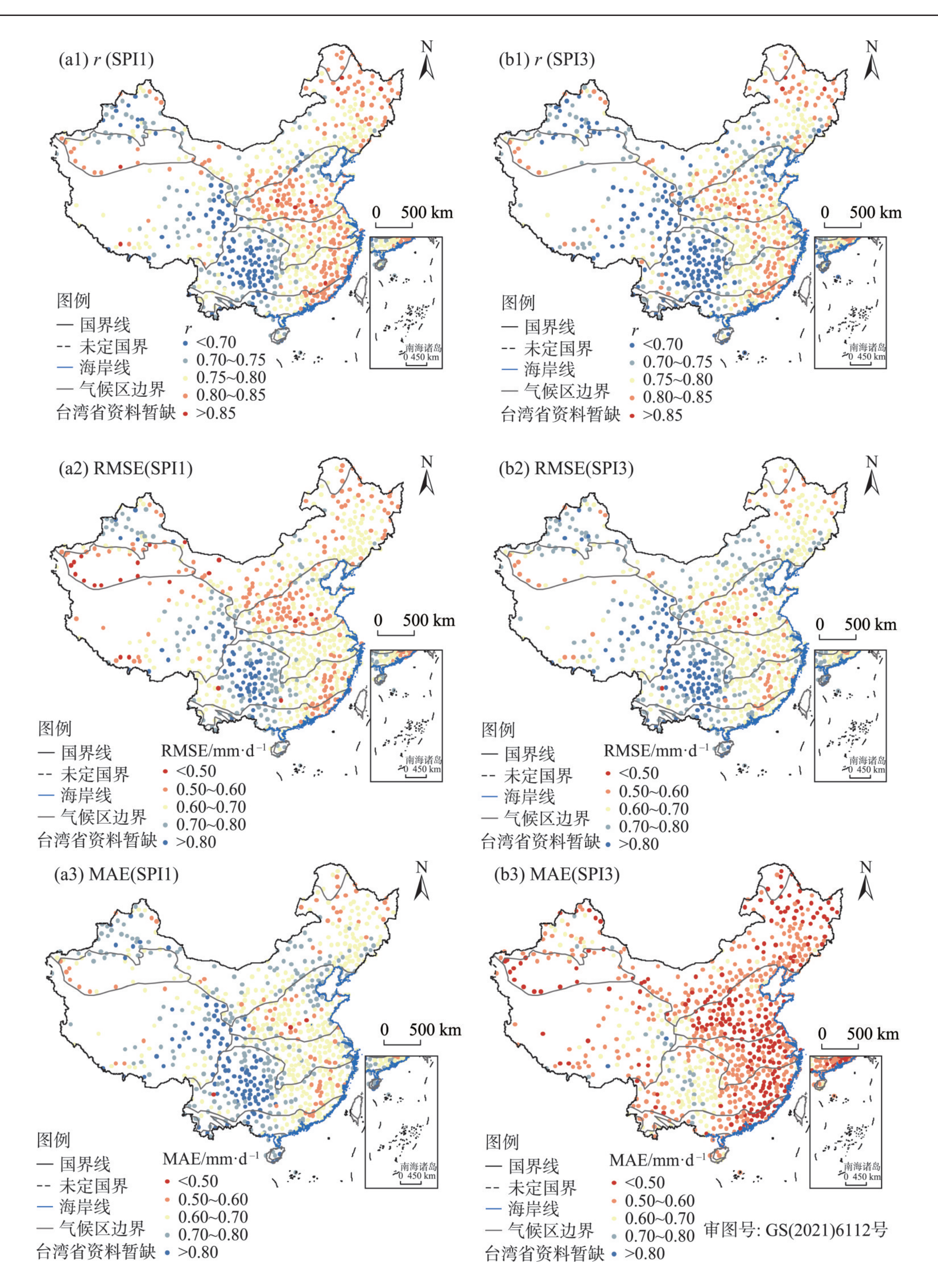


图6 ERA5计算的SPI1和SPI3的精度

Fig. 6 Accuracy of SPI1 and SPI3 calculated by ERA5

精度低，这与以往对于再分析降水数据的研究结果[16,28-30]类似。ERA5数据在不同区域的精度存在差异的原因有很多，首先ERA5为再分析数据，是使用多种资料数据通过同化而来的，其精度受原始输入数据资料质量和同化算法的影响[15]；在观测数据稀疏的地方或输入数据质量低的区域再分析产品的数据质量也相对低；其次，降水具有较强的局域性，而ERA5数据为栅格数据，两者之间存在一定的尺度问题，用站点所在的ERA5格点的降水值表示站点的降水会和站点真实观测的降水存在一定偏差。

此外，本文的研究还存在一些不足。首先，全国的气象站点分布不均，导致不同气候区和不同海拔梯度区内的站点分布不均、数量不同。大部分站

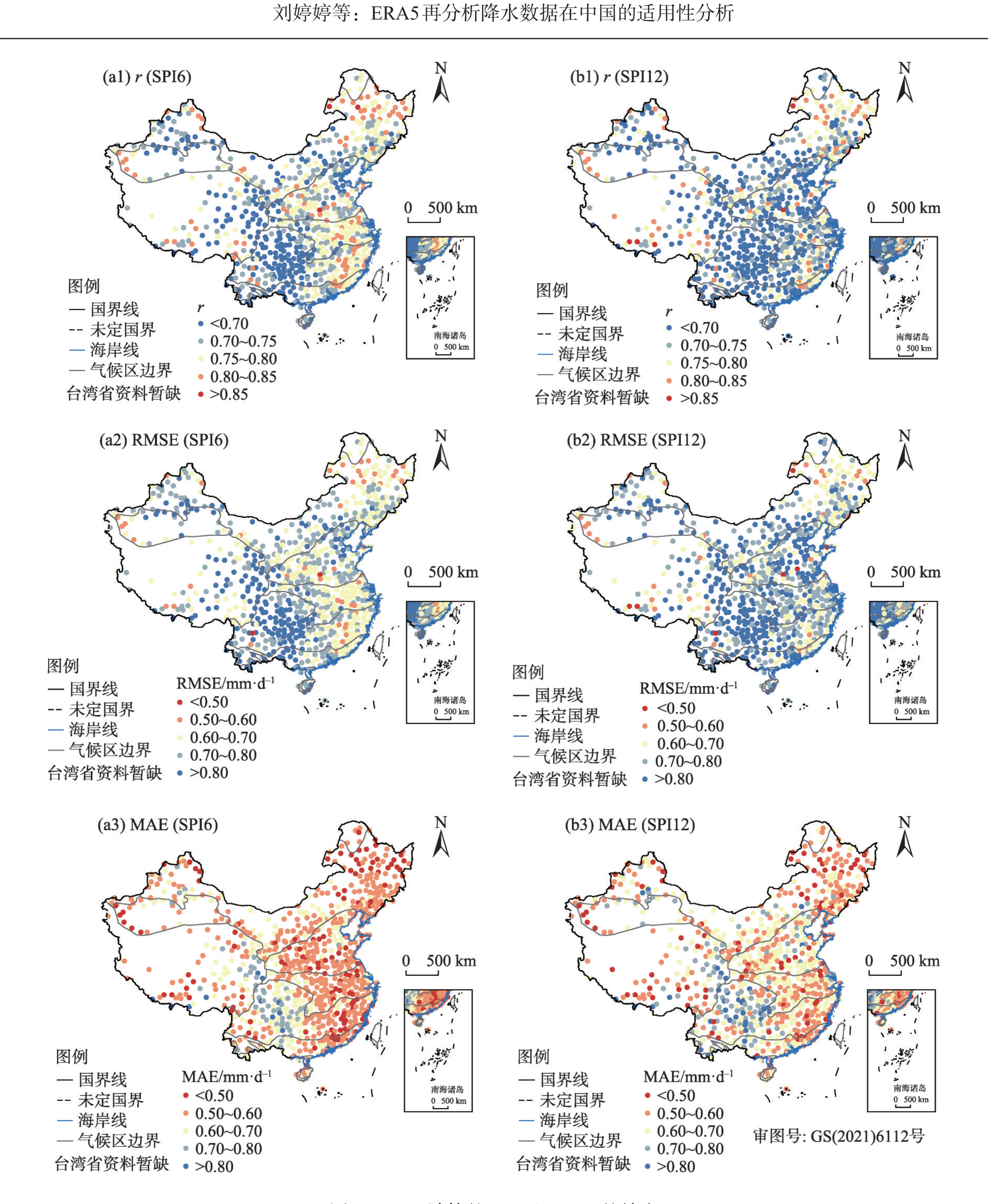


图7 ERA5计算的SPI6和SPI12的精度

Fig. 7 Accuracy of SPI6 and SPI12 calculated by ERA5

点分布在中国东部区域和低海拔地区。对于站点少的区域,验证结果可能会存在一定偏差。其次,ERA5数据的空间分辨率为0.25°,反映的0.25°×0.25°格网内区域降水的平均情况,站点尺度的降水反映的是该站点一定范围内的降水情况,然而降水在局部地区(特别是地形变化复杂)的变化可能会很大,直接对两者进行对比也会产生一定误差。

4 结 论

本文以站点观测的降水数据为参考,分析了ERA5再分析降水数据在不同时间尺度(月、季)、不同气候区、不同海拔梯度下的精度以及ERA5再分析数据对暴雨和干旱事件的刻画能力。得到主要结论如下:

表3 阈值设置为-0.5时基于ERA5识别的干旱在各气候区POD、FAR和ETS的均值

Tab. 3 Mean values of POD, FAR and ETS in each climate zone with threshold of −0.5 for drought identification

气候区	探测率(POD)	误报率(FAR)	公平先兆评分(ETS)
高原气候区	0.653	0.334	0.351
北温带	0.717	0.291	0.419
中温带	0.685	0.302	0.386
南温带	0.693	0.295	0.401
北亚热带	0.706	0.298	0.403
中亚热带	0.661	0.327	0.356
南亚热带	0.688	0.305	0.388
北热带	0.662	0.326	0.352

(1) ERA5降水数据对日降水事件的识别能力在空间和时间上均存在差异。8个气候带中,精度最高的是北温带,其日降水和站点观测的r、RMSE和MAE分别为0.587、4.040 mm·d^{-1}和1.472 mm·d^{-1}。海拔>500 m地区的精度低于海拔≤500 m地区的精度,冬、春两季较夏、秋两季的精度高。

(2) ERA5数据在进行暴雨的识别时,与站点数据的偏差较大,且阈值越大(即暴雨越强)偏差越大,整体来看对东南部(北亚热带、中亚热带东部和南亚热带东部)的暴雨事件捕捉的最好。

(3) 基于ERA5计算的月(1个月)、季(3个月)、

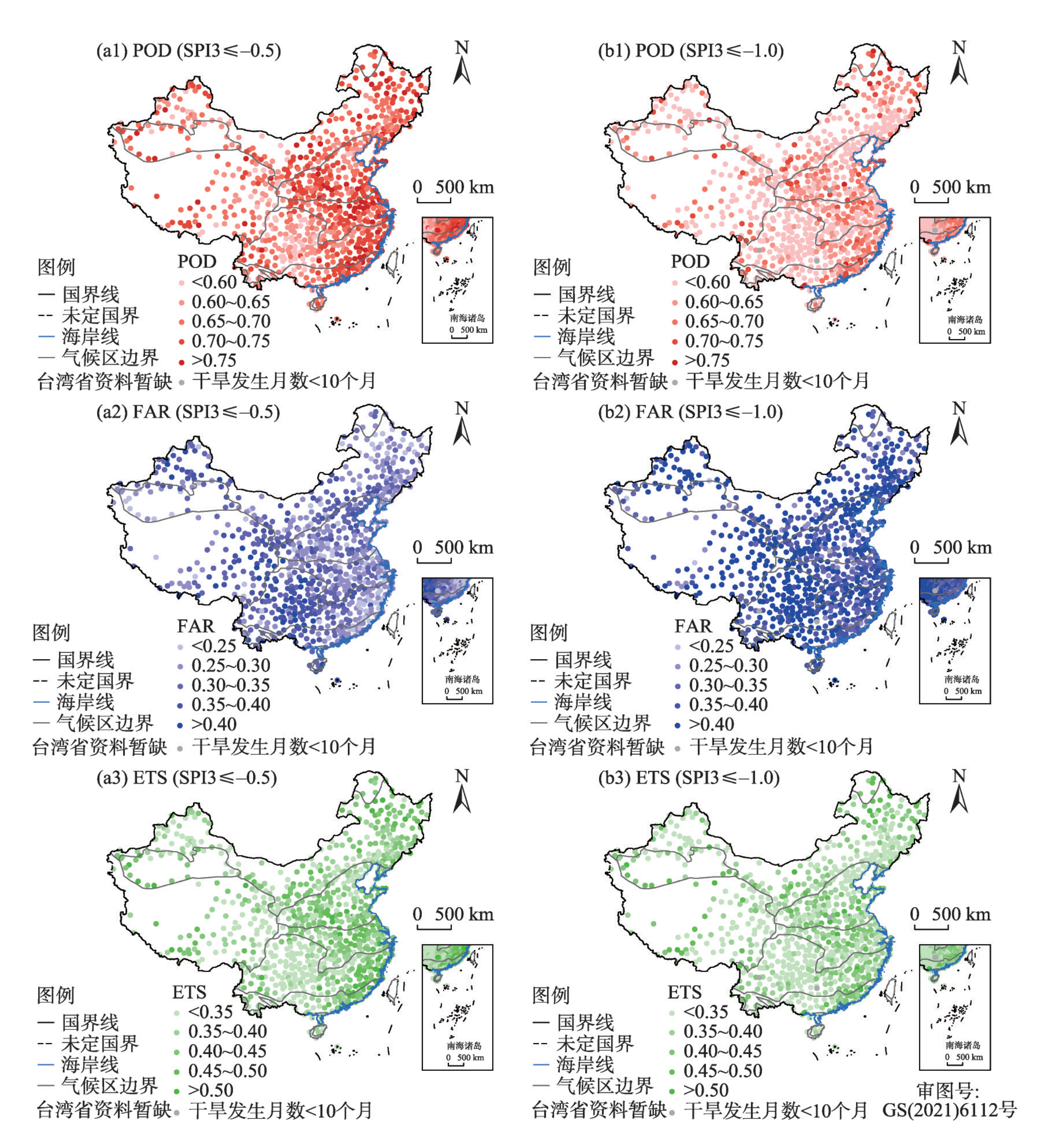


图8 ERA5识别阈值为-0.5和-1.0时干旱的精度

Fig. 8 Accuracy of ERA5 in recognizing drought with the thresholds of −0.5 and −1.0

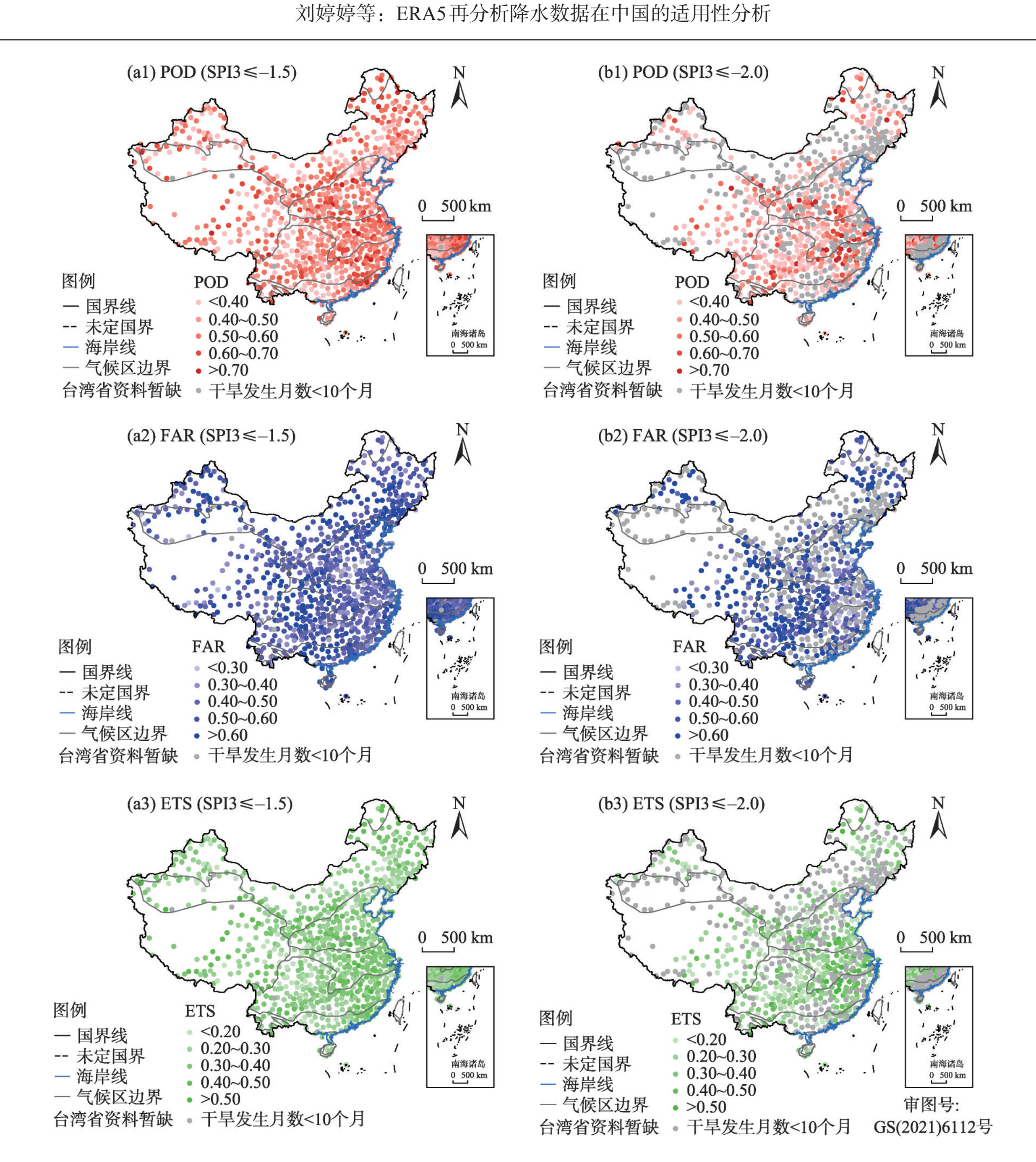


图9 ERA5识别阈值为-1.5和-2.0时干旱的精度

Fig. 9 Accuracy of ERA5 in recognizing drought with thresholds of -1.5 and -2.0

半年(6个月)和年(12个月)尺度的SPI的精度不同，其中SPI3的精度最高。在进行干旱事件的识别时，与站点数据有较大的差距，且阈值越低(即干旱越严重)误差越大，整体来看对北温带、南温带和北亚热带的干旱捕捉能力较好。

参考文献(References)

[1] 夏军, 谈戈. 全球变化与水文科学新的进展与挑战[J]. 资源科学, 2002, 24(3): 1-7. [Xia Jun, Tan Ge. Hydrological science towards global change: Progress and challenge[J]. Resources Science, 2002, 24(3): 1-7.]

[2] Adler R F, Huffman G J, Chang A, et al. The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979- present) [J]. Journal of hydrometeorology, 2003, 4(6): 1147-1167.

[3] 谈戈, 夏军, 李新. 无资料地区水文预报研究的方法与出路[J]. 冰川冻土, 2004, 26(2): 192-196. [Tan Ge, Xia Jun, Li Xin. Hydrological prediction in ungauged basins[J]. Journal of Glaciology and Geocryology, 2004, 26(2): 192-196.]

[4] Su F, Hong Y, Lettenmaier D P. Evaluation of TRMM multisatellite precipitation analysis (TMPA) and its utility in hydrologic prediction in the La Plata Basin[J]. Journal of Hydrometeorology,

2008, 9(4): 622-640.

[5] 黄建平, 张国龙, 于海鹏, 等. 黄河流域近40年气候变化的时空特征[J]. 水利学报, 2020, 51(9): 1048-1058. [Huang Jianping, Zhang Guolong, Yu Haipeng, et al. Characteristics of climate change in the Yellow River Basin during recent 40 years[J]. Journal of Hydraulic Engineering, 2020, 51(9): 1048-1058.]

[6] 黄颖, 毛文茜, 王潇雅, 等. 近39 a祁连山及其周边地区降水量时空分布特征[J]. 干旱气象, 2020, 38(4): 527-534. [Huang Ying, Mao Wenqian, Wang Xiaoya, et al. Temporal and spatial distribution of precipitation in the Qilian Mountain and its surrounding areas in recent 39 years[J]. Journal of Arid Meteorology, 2020, 38(4): 527-534.]

[7] 赵建婷, 王艳君, 苏布达, 等. 印度河流域气温、降水、蒸发及干旱变化特征[J]. 干旱区地理, 2020, 43(2): 72-82. [Zhao Jianting, Wang Yanjun, Su Buda, et al. Spatiotemporal distributions of temperature, precipitation, evapotranspiration, and drought in the Indus River Basin[J]. Arid Land Geography, 2020, 43(2): 72-82.]

[8] 徐昆, 朱秀芳, 刘莹, 等. 采用AquaCrop作物生长模型研究中国玉米干旱脆弱性[J]. 农业工程学报, 2020, 36(1): 154-161. [Xu Kun, Zhu Xiufang, Liu Ying, et al. Vulnerability of drought disaster of maize in China based on AquaCrop model[J]. Transactions of the Chinese Society of Agricultural Engineering, 2020, 36(1): 154-161.]

[9] 徐昆, 朱秀芳, 刘莹, 等. 气候变化下干旱对中国玉米产量的影响[J]. 农业工程学报, 2020, 36(11): 149-158. [Xu Kun, Zhu Xiufang, Liu Ying, et al. Effects of drought on maize yield under climate change in China[J]. Transactions of the Chinese Society of Agricultural Engineering, 2020, 36(11): 149-158.]

[10] 张小丽, 彭勇, 王本德, 等. 基于SWAT模型的降雨数据适用性评价[J]. 农业工程学报, 2014, 30(19): 88-96. [Zhang Xiaoli, Peng Yong, Wang Bende, et al. Suitability evaluation of precipitation data using SWAT model[J]. Transactions of the Chinese Society of Agricultural Engineering, 2014, 30(19): 88-96.]

[11] 焦振航, 舒红, 吴凯, 等. 降水驱动数据改进对VIC土壤湿度模拟的影响[J]. 城市勘测, 2017(4): 37-41. [Jiao Zhenhang, Shu Hong, Wu Kai, et al. The rainfall calibration methods' impact on VIC soil moisture simulation[J]. Urban Surveying, 2017(4): 37-41.]

[12] Tarek M, Brissette F P, Arsenault R. Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modelling over North America[J]. Hydrology and Earth System Sciences, 2020, 24(5): 2527-2544.

[13] 冯克鹏, 洪阳, 田军仓, 等. 多源降水数据的小流域水文模拟效用评估[J]. 干旱区地理, 2020, 43(5): 1179-1191. [Feng Kepeng, Hong Yang, Tian Juncang, et al. Evaluating runoff simulation of multi-source precipitation data in small watersheds[J]. Arid Land Geography, 2020, 43(5): 1179-1191.]

[14] Albergel C, Dutra E, Munier S, et al. ERA-5 and ERA-Interim driven ISBA land surface model simulations: Which one performs better?[J]. Hydrology and Earth System Sciences, 2018, 22(6): 3515-3532.

[15] 韦芬芬, 汤剑平, 王淑瑜. 中国区域夏季再分析资料高空变量可信度的检验[J]. 地球物理学报, 2015, 58(2): 383-397. [Wei Fenfen, Tang Jianping, Wang Shuyu. A reliability assessment of upper-level reanalysis datasets over China[J]. Chinese Journal of Geophysics, 2015, 58(2): 383-397.]

[16] 胡增运, 倪勇勇, 邵华, 等. CFSR, ERA-Interim和MERRA降水资料在中亚地区的适用性[J]. 干旱区地理, 2013, 36(4): 700-708. [Hu Zengyun, Ni Yongyong, Shao Hua, et al. Applicability study of CFSR, ERA-Interim and MERRA precipitation estimates in Central Asia[J]. Arid Land Geography, 2013, 36(4): 700-708.]

[17] Hersbach H, Bell B, Berrisford P, et al. The ERA5 global reanalysis[J]. Quarterly Journal of the Royal Meteorological Society, 2020, 146(730): 1999-2049.

[18] Graham R M, Hudson S R, Maturilli M. Improved performance of ERA5 in Arctic gateway relative to four global atmospheric reanalyses[J]. Geophysical Research Letters, 2019, 46(11): 6138-6147.

[19] Hénin R, Liberato M L, Ramos A M, et al. Assessing the use of satellite-based estimates and high-resolution precipitation datasets for the study of extreme precipitation events over the Iberian Peninsula[J]. Water, 2018, 10(11): 1688, doi: 10.3390/w10111688.

[20] Wang C, Graham R M, Wang K, et al. Comparison of ERA5 and ERA-Interim near-surface air temperature, snowfall and precipitation over Arctic sea ice: Effects on sea ice thermodynamics and evolution[J]. The Cryosphere, 2019, 13(6): 1661-1679.

[21] Betts A K, Chan D Z, Desjardins R L. Near-surface biases in ERA5 over the Canadian prairies[J]. Frontiers in Environmental Science, 2019, 7: 129, doi: 10.3389/fenvs.2019.00129.

[22] Beck H E, Pan M, Roy T, et al. Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS[J]. Hydrology and Earth System Sciences, 2019, 23(1): 207-224.

[23] Nogueira M. Inter-comparison of ERA-5, ERA-interim and GPCP rainfall over the last 40 years: Process-based analysis of systematic and random differences[J]. Journal of Hydrology, 2020, 583: 124632, doi: 10.1016/j.jhydrol.2020.124632.

[24] Xu X, Frey S K, Boluwade A, et al. Evaluation of variability among different precipitation products in the Northern Great Plains[J]. Journal of Hydrology: Regional Studies, 2019, 24: 100608, doi: 10.1016/j.ejrh.2019.100608.

[25] Amjad M, Yilmaz M T, Yucel I, et al. Performance evaluation of satellite-and model-based precipitation products over varying climate and complex topography[J]. Journal of Hydrology, 2020, 584: 124707, doi: 10.1016/j.jhydrol.2020.124632.

[26] Fallah A, Rakhshandehroo G R, Berg P, et al. Evaluation of precipitation datasets against local observations in southwestern Iran[J]. International Journal of Climatology, 2020, 40(9): 4102-4116.

[27] Jiang Q, Li W, Fan Z, et al. Evaluation of the ERA5 reanalysis precipitation dataset over Chinese Mainland[J]. Journal of Hydrology, 2020: 125660, doi: 10.1016/j.jhydrol.2020.125660.

[28] 成晓裕, 王艳华, 李国春, 等. 三套再分析降水资料在中国区域的对比评估[J]. 气候变化研究进展, 2013, 9(4): 258-265. [Cheng Xiaoyu, Wang Yanhua, Li Guochun, et al. Evaluation of three reanalysis precipitation datasets in China[J]. Climate Change Research, 2013, 9(4): 258-265.]

[29] 孙葭, 章新平, 黄一民. 不同再分析降水数据在洞庭湖流域的精度评估[J]. 长江流域资源与环境, 2015, 24(11): 1850–1859. [Sun Jia, Zhang Xinping, Huang Yimin. Evaluation of precipitation from ERA-Interim, CRU, GPCP and TRMM reanalysis data in the Dongting Lake Basin[J]. Resources and Environment in the Yangtze Basin, 2015, 24(11): 1850–1859.]

[30] 刘鹏飞, 刘丹丹, 梁丰, 等. 三套再分析降水资料在东北地区的适用性评价[J]. 水土保持研究, 2018, 25(4): 215–221. [Liu Pengfei, Liu Dandan, Liang Feng, et al. Comparison the adaptability of CFSR, MERRA, NCEP reanalysis precipitation data and observation in northeast China[J]. Research of Soil and Water Conservation, 2018, 25(4): 215–221.]

Applicability of ERA5 reanalysis of precipitation data in China

LIU Tingting[1,2], ZHU Xiufang[1,2,3], GUO Rui[1,2], XU Kun[1,2], ZHANG Shizhe[1,2]

(1. Key Laboratory of Environmental Change and Natural Disaster of Ministry of Education, Beijing Normal University, Beijing 100875, China; 2. Institute of Remote Sensing Science and Engineering, Faculty of Geographic Sciences, Beijing Normal University, Beijing 100875, China; 3. State Key Laboratory of Remote Sensing Science, Beijing Normal University, Beijing 100875, China)

Abstract: ERA5 is the latest generation of ECMWF reanalysis product. Its precipitation data can be used to analyze the spatial and temporal characteristics of precipitation, extracting rainstorm and flood events, driving crop models, hydrological models, and land surface models. Studies on the effects of using reanalysis data to drive various models to simulate climate change are becoming more common in the context of climate change. However, as atmospheric reanalysis data, the data quality of ERA5 precipitation data will be affected by the errors of forecast products, observation data, and assimilation methods. The precision of reanalysis data will influence the uncertainty of related study results. As a result, verifying and evaluating the accuracy of reanalysis data is both necessary and urgent. To investigate the applicability of ERA5 precipitation data in China, daily precipitation data from 728 Chinese sites were used as a reference to examine the accuracy of ERA5 reanalysis precipitation data at various time scales (month, quarter), different climate zones and different elevation gradients using the Pearson correlation coefficient (r), root mean square error (RMSE), mean absolute error (MAE), probability of detection (POD), false alarm rate, and equitable threat score. Furthermore, the capability of ERA5 to depict heavy rain and drought events was investigated. The results demonstrate that there are spatial and temporal differences in the ability of ERA5 precipitation data to identify daily precipitation events. Among the eight climatic zones, the north temperate zone has the highest accuracy. The r, RMSE, and MAE of ERA5 daily precipitation and station observation are 0.587, 4.040 mm·d^{-1}, and 1.472 mm·d^{-1}, respectively, in the north temperate zone. Precipitation data from the ERA5 are less accurate in the summer and autumn than in the winter and spring. The accuracy of ERA5 precipitation data in areas with elevations greater than >500 m is lower than in areas with elevations of less than ≤500 m. When the ERA5 data is used to identify the heavy rain, there is a large deviation from observations of the stations, and the larger the threshold (i.e., the stronger the heave rain), the larger the deviation is. The accuracy of standardized drought index (SPI) of different time scales calculated using ERA5 is different, varies with SPI on a three-month time scale having the highest accuracy. When using ERA5 data to identify drought events, the lower the threshold (that is, the greater the severity of the drought), the greater the error. Generally speaking, the drought capture ability of the north temperate zone, south temperate zone, and north subtropical zone is better. This research broadens the scope of validation of climate data sets and serves as a model for future research. These findings serve as a reference for related studies considering whether to use ERA5 data, as well as a tool for analyzing the uncertainties of related studies.

Key words: ERA5 data; precipitation; heavy rain; drought